Chapter 3
Natural Goodness

Dennis L. Krebs

Abstract Although the theory of evolution might seem to imply that all animals must be selfish by nature, this is not the case because social animals can propagate their selfish genes in psychologically unselfish ways. To understand the moral aspect of human nature, we must understand the adaptive functions that moral traits served in early human environments. I argue that the central function of morality is to uphold adaptive systems of cooperation. Even though some cooperative strategies are susceptible to exploitation by selfish strategies, there are several ways in which the kind of cooperative behavioral strategies that people consider moral can evolve. Primitive psychological sources of moral behavior, such as moral emotions, and advanced sources, such as perspective-taking and moral reasoning, evolve and develop throughout the life span in a Russian Doll manner. Although the original function of perspective-taking and moral reasoning may have been to help early humans advance their interests in strategic social interactions, these processes may now motivate people to behave in moral ways. We are evolved to be as good as our early ancestors had to be to reap the benefits of sociality and cooperation.

3.1 Introduction

Some eminent evolutionary biologists have asserted that, on the laws of evolution, all animals must be selfish by nature. For example, George Williams (1989) wrote, "there is no encouragement for any belief that an organism can be designed for any purpose other than the most effective pursuit of … self-interest…. [because] evolution is guided by a force that maximizes genetic selfishness" (pp. 195–6). In a similar vein, in *The Selfish Gene*, Dawkins (1989) asserted, "'Nature red in tooth and claw' sums up our modern understanding of natural selection admirably…. Much as we might wish to believe otherwise, universal love and welfare of the species as a whole are concepts that simply do not make evolutionary sense …. Be warned that

D. L. Krebs (✉)
Simon Fraser University, Burnaby, BC, Canada
e-mail: dennis_krebs@sfu.ca

J. van Nes et al. (eds.), *Relational Anthropology for Contemporary Economics*,
Ethical Economy 61, https://doi.org/10.1007/978-3-030-84690-9_3

if you wish, as I do, to build a society in which individuals cooperate generously and unselfishly toward a common good, you can expect little help from biological nature" (pp. 3–4).

Claims such as these stem from the assumption that it is appropriate to characterize those who win "survival of the fittest" contests as selfish because they propagate their genes at the expense of those against whom they are competing. It would seem that a mutation that disposed an animal to sacrifice its prospects of surviving and reproducing for the sake of others would not fare well in the process of natural selection. For this reason, so the argument goes, on the laws of evolution, all animals must be selfish by nature.

There is, however, an obvious problem with this conclusion, namely that many animals appear to behave in unselfish ways. Members of many species make sacrifices for their offspring and other relatives. Members of some species join forces to kill prey, and members of other species help one another avoid predators, by for example emitting alarm calls. Many animals collaborate in the building of shelters, dams, and tunnels. Although most people assume that we are the only – or at least the most – altruistic species, we do not even come close to the sacrifices that many social insect species make for their groups. Ants, bees, and termites spend virtually all of their lives gathering food for others, and they willingly sacrifice their lives to defend their colonies.

If biologists such as George Williams and Richard Dawkins were correct in asserting that all traits that evolve through natural selection must be selfish, then evidence that animals behave in altruistic and cooperative ways would constitute a serious challenge to the theory of evolution. Charles Darwin was aware of this challenge, acknowledging that the self-sacrificial altruism of social insects presented "one special difficulty, which at first seemed insuperable, and actually fatal to my whole theory." In a similar vein, Williams (1989) asked, "how could maximizing selfishness produce an organism capable of often advocating, and occasionally practicing, charity toward strangers and even towards animals" (p. 208)?

One of the ways in which skeptical evolutionary theorists have accounted for seemingly altruistic behaviors will seem familiar to many – they have attributed it to social learning and cultural indoctrination. As expressed by Dawkins (1989), "let us try to *teach* generosity and altruism, because we are born selfish…. Our genes may instruct us to be selfish, but we are not necessarily compelled to obey them all our lives" (p. 3). Or, as expressed by Campbell (1978), "in man, genetic competition precludes the evolution of … genetic altruism. The behavioral dispositions that produce … self-sacrificial altruism must instead be products of culturally evolved indoctrination that has had to counter self-serving genetic tendencies." According to Campbell, an important implication of this point is that "man is profoundly ambivalent in his social role – as Freud noted…. The commandments, the proverbs, the religious 'law' represent social evolutionary products directed at inculcating tendencies that are in direct opposition to the 'temptations' representing, for the most part, the dispositional tendencies produced by biological evolution" (pp. 52–3). Campbell and Dawkins are endorsing Original Sin-type models of human nature. Many people find such models appealing: we are born bad, but we can learn to be good.

There is no question that skeptical evolutionary theorists are correct when they assert that we are evolved to behave in sinful, selfish, and immoral ways, and there is no question that they are correct when they assert that we can be taught to be good. However, this is just half the story. We also are evolved to behave in unselfish and moral ways, and we can learn to be bad. In essence, Original Sin models pit nature against nurture, biology against culture, biological evolution against cultural evolution, and genes against environments. However, biology and culture can work together to produce moral traits. I believe that the evidence supports the conclusion that mental mechanisms that dispose us to behave in ways that we consider altruistic, fair, and moral can evolve and have evolved in our species (and in some other species as well) (Krebs 2011). We inherit a capacity to be good. This does not mean that we are entirely good by nature. We are evolved to be good and bad, moral and immoral, virtuous and vicious, depending on the conditions.

3.2 The Nature of Selfishness and Altruism

The first step in explaining how a capacity to be good could evolve is to recognize that genetic selfishness is quite different from the kind of selfishness that we consider bad. The kind of selfishness that is relevant to morality does not pertain to the *propagation of genes;* it pertains to the *motives of individuals*. Several social scientists have pointed out that a great deal of confusion and misunderstanding about human nature has been fomented by scholars who fail to recognize that biologists define selfishness and altruism in qualitatively different ways from the ways in which psychologists and laypeople define these constructs. For example, in response to an assertion by Richard Dawkins that genes that rendered horses susceptible to developing bad teeth and genes that disposed humans to smoke cigarettes would qualify as altruistic if they reduced individuals' prospects of surviving and reproducing, Batson (2000) wrote, "most people interested in the existence of altruism are not thinking about bad teeth in horses or smoking cigarettes; they are thinking about psychological altruism" (p. 208). In a related vein, Sober and Wilson (2000) pointed out that "the automatic assumption that individualism [i.e., selfishness] in evolutionary biology and egoism in the social sciences must reinforce each other is as common as it is mistaken. More care is needed to connect the behaviors that evolved … with the psychological mechanisms that evolved to motivate those behaviors" (p. 205). Expanding on this point, De Waal (2008) explained that although evolutionary accounts of altruism are "built around the principle that all that natural selection can work with are the effects of behavior, not the motivation behind it," they persist in invoking motivational terms:

> The hijacking of motivational terminology by evolutionary biologists has been unhelpful for communication about motivation per se…. It is not for nothing that biologists hammer on the distinction between ultimate and proximate ….ultimate accounts stress return-benefits, i.e., positive consequences for the performer and/or its kin. Inasmuch as these benefits may be quite delayed, however, it is unclear what motivational role, if any, they play (pp. 280–1).

Whether selfish individuals who seek to obtain benefits for themselves without due regard for others fare better biologically and contribute more copies of their genes to future generations than those who behave in more moral ways is an open question. They might, or they might not. There is nothing in the process of natural selection that dictates that individuals who are motivated to behave in ways that we consider bad will prevail in the struggle for existence. Psychologically altruistic individuals motivated to help others as an end in itself, and moral individuals motivated to benefit themselves and others in fair and equitable ways, could be more likely to survive, to produce offspring, and to propagate their genes than individuals who are motivated to advance their own interests without concern for others. Altruistic motives could produce biologically beneficial results. People who genuinely want to help others could fare better biologically than people who are concerned only with themselves. The assumption that all evolved dispositions are selfish is valid only with respect to genetic forms of selfishness in the environments in which they were selected. Inasmuch as moral traits can be genetically selfish (i.e., can increase the biological success of those who emit them), they can evolve.

The question that those who are concerned with human nature should be asking is how *psychologically selfish* strategies, defined in terms of the motivation to advance one's own welfare without concern for others, fared against *psychologically unselfish* strategies in helping early humans propagate their genes (that is to say, achieve genetically selfish effects). Which strategies were selected and evolved to become part of human nature? Did those who were motivated to cooperate, behave fairly, and help others contribute more copies of their genes to future generations than those who were motivated to look out only for themselves?

3.3 The Evolution of Social Strategies

Imagine members of early human groups faced with recurring decisions about whether to behave in selfish or unselfish ways. Evolutionary theorists assume that such choices are guided by genetically influenced strategies, and they seek to understand which strategies produced the most adaptive decisions in early human environments and, therefore, which ones evolved. Although the genes that program evolved strategies were selected hundreds of thousands, even millions, of years ago, evolutionary theorists expect them to guide strategic decision making in modern environments in conditions corresponding to those that regulated them in archaic environments.

3.3.1 The Adaptive Potential in Cooperative Strategies

It is easy to see how animals could advance their adaptive interests more effectively by adopting cooperative social strategies than by adopting solitary or selfish strategies. Examples abound in the animal kingdom. We can safely assume that early humans were poorly equipped to survive on their own. A solitary human with small teeth, fingernail claws, devoid of fur to keep warm, and relatively slow afoot would not have lasted long in early environments, just as modern humans would not last long by themselves in the wild today. Although our distant primate ancestors possessed killer ape traits such as large canine teeth and claws, these traits diminished as we evolved because early humans acquired adaptations that enabled them to solve their adaptive problems more effectively in social ways. As expressed by Curry (2016), "humans descended from a long line of social primates; they have spent 50 million years living in social groups…. and two million years making a living as intensely collaborative hunter-gatherers. This has equipped humans with a range of biological – including psychological – adaptations for cooperation" (p. 29). It is easy to see how early humans who inherited genes that induced them to join forces to kill large game and defend their groups, to collaborate in the construction of shelters, to help one another when they were in need, and to coordinate their efforts to mate and rear offspring could have fared better than early humans who inherited genes that induced them to behave in more selfish and immoral ways (Ellis 1998).

There is untold adaptive potential in cooperative social strategies in modern societies. If everyone did his or her share, we would produce significantly more resources than we currently produce, and if everyone took his or her share, we would not have to waste energy competing for them. If no one were disposed to cheat, we wouldn't have to waste resources on crime prevention and the punishment of criminals. Everyone would benefit. So, in view of the tremendous adaptive potential in cooperation, why aren't we more cooperative and moral?

3.3.2 Obstacles to the Evolution of Cooperation: The Adaptive Potential in Selfish Strategies

The reason that unconditionally cooperative (purely moral) strategies have not evolved in the human species is because it is in the biological and genetic interest of individuals to maximize their gains and to minimize their costs in exchanges with others, and this creates the temptation to behave selfishly and cheat. As expressed by Rawls (1999) in the opening pages of his classic book, *A Theory of Justice*:

> Although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts. There is a conflict of interests since persons are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share (p. 4).

Conflicts of interest give rise to moral problems because they tempt us to advance our interests at the expense of others, which undermines cooperative social orders. In the words of Alexander (1985), "if there were not conflicts of interest among people and societies it is difficult to see how concepts of right and wrong, ethics and morality, and selfishness and altruism could ever have arisen." Alexander argues that to understand conflicts of interest fully, we must trace them back to their biological core: "The interests of every individual human (i.e., the directions of its striving) are expected to be toward ensuring the indefinite survival of its genes and their copies, whether these are resident in the individual, its descendants, or its collateral relatives" (p. 3).

Conflicts of interest present obstacles to the evolution of cooperative and fair strategies. For example, it would have been in the interest of early humans who collaborated in hunting large game and defending themselves against predators to conserve their energy, position themselves in ways that diminished their chances of getting injured, let others do the dirty work, and take the lion's share of the spoils. Such cheating is prevalent in the animal kingdom and among modern humans in some situations.

The biological benefits of strategies inducing individuals to behave in cooperative and moral ways hinge on them interacting with other individuals who behave in cooperative and moral ways. The problem is, within populations of individuals inheriting genes that induce them to behave in unconditionally cooperative ways, some individuals inevitably would inherit genes that disposed them to behave in more selfish ways, and unfortunately, unless counteracted effectively, the selfish individuals would come out ahead. The bad guys would take what that the good guys were willing to offer without suffering the costs of giving in return, and the good guys would suffer the costs of giving to the bad guys without reaping the benefits of receiving in return. In the currency of evolution, the selfish individuals would be more likely than the cooperative individuals to survive and propagate selfish offspring like themselves, who would be more likely to propagate additional selfish offspring, and so on, causing an exponential explosion of selfish individuals in the population. It follows that strategies that induce individuals to behave in *unconditionally* cooperative and moral ways are doomed to extinction in groups containing individuals who are willing and able to exploit cooperative individuals by behaving in selfish ways. The only population in which unconditionally cooperative strategies could evolve would be one in which there were no genetic conflicts of interest, such as in a population of clones. In sexually reproducing species such as our own, pure goodness is out of the question.

3.3.3 The Ultimate Irony: Self-Defeating Selfishness

In view of the adaptive superiority of social strategies that induce individuals to behave in selfish ways when competing against strategies that induce them to behave in unconditionally cooperative ways, it might seem that selfish strategies would win all evolutionary contests, rendering all species selfish by nature, as skeptical evolutionary theorists, such as those I quoted above, have claimed. Fortunately, however, this is not the case. Even though strategies that induce individuals to behave in selfish ways can evolve and have evolved in many species, including our own, unconditionally selfish strategies are not optimal in social species that need, or can benefit from, assistance from others. Brown (1984) eloquently explained why.

Imagining a group of cooperative Christian birds dependent for their survival on being groomed by other birds, Brown (1984) considered the ultimate effect of a mutant cheater who reaped the benefits of being groomed without suffering the costs of grooming others in return. As discussed, the cheater would fare better than the cooperative members of the group, propagating offspring who inherited his or her cheating ways, and so on, until the selfish cheaters replaced all of the Christian cooperators. However, Brown points out that this would usher in a tragically self-defeating consequence for the selfish cheaters, because "once grooming birds had become extinct, so eventually would cheaters; one imagines a pathetic final act in which all birds on the stage present to one another heads that none will groom" (Brown 1984). Although it might have been in the biological interest of members of early human groups that collaborated in activities such as hunting prey and defending themselves against predators to do less than their share and take more than their share under some circumstances, this strategy would not have paid off if other members also adopted it or if they got punished for shirking their duties.

The evolution of unconditionally selfish and immoral strategies would be an unmitigated disaster in the human species. No one would help anyone. Everyone would try to cheat everyone else. Unconditional selfishness would inevitably do everyone down because in social species in which individuals are dependent on one another for their welfare, those on whom they are dependent – whether marital partners, friends, or members of their groups – are resources that it is in their interest to preserve and cultivate. Helping those on whom your welfare is dependent is like cultivating resources such as gardens, orchards, and domesticated animals that you can use to advance your welfare down the line. Investing in your group is like investing in public goods such as the common grazing land described by Garrett Hardin in *The Tragedy of the Commons*. Failing to support those on whom you are dependent is like letting your resources go to waste. As expressed by Flack and de Waal (2000):

> Inasmuch as every member [of a group] benefits from a unified, cooperative group, one expects them to care about the society they live in, and to make an effort to improve and strengthen it similar to the way the spider repairs her web, and the beaver maintains the integrity of his dam. Each and every individual has a stake in the quality of the social environment on which its survival depends. In trying to improve this quality for their own purposes, they help many of their group mates at the same time (p. 14).

To summarize, we would expect individuals who inherited genes that disposed them to adopt cooperative moral strategies to fare better than individuals who inherited genes that disposed them to adopt selfish immoral strategies as long as the good guys interacted with other good guys and the bad guys interacted with other bad guys. In addition, we would expect cooperative groups to prevail in competitions against selfish groups – a phenomenon that some theorists believe played a very important role in the evolution of morality. However, within mixed groups, we would expect selfish strategists who exploit cooperative strategists to come out ahead, but only temporarily. If selfish strategists ran cooperative strategists into the ground, they would destroy the resources they needed to advance their welfare, leaving them with no choice but to go it alone or to interact with one another. The solution to this problem – the way in which natural selection solved the problem of selfishness – is to build conditions into cooperative strategies that induce co-operators to interact with co-operators and guard against being exploited by those who are prone to behave selfishly.

3.3.4 Resolving the Conundrum of Cooperation: The Evolution of Conditional Strategies

Evolutionary game theorists have created computer-based simulations of social evolution in which they have pitted a variety of social strategies against one another. In these contests, points equate to replicas, or genetic clones, of winning strategies. The better each strategy does against the other strategies against which it competes, the more copies of itself it contributes to future generations, and the more copies of defeated strategies it knocks out of the population.

"Behave selfishly" is a powerful social strategy – probably the most primitive of all. It doesn't require much brainpower. It is uncomplicated, and it is a daunting foe in social games. One on one, it can't be beaten in single exchanges; the best that other strategies can do is tie it. Pitted against cooperative strategies, behaving selfishly always wins the initial exchange. In one-shot interactions, the best you can do against a selfish opponent is make a selfish move yourself, and tie. For reasons such as these, many evolutionary game theorists assumed that there were no strategies that could defeat purely selfish strategies (and therefore that all animals were selfish by nature), but this changed when some game theorists found that a simple conditionally cooperative strategy, Tit for Tat, could prevail over more selfish strategies in repeated games and evolve under certain conditions.

Tit for Tat contains the decision-rule "on the initial exchange with other players, make a cooperative move, then on the following exchanges, treat them the way they treated you." "If your partner is nice to you, be nice back, but if your partner treats you selfishly, treat him or her selfishly in return." This decision-rule directs players to make a cooperative overture to potential exchange partners and follow it up by reciprocating their decisions. Although the Tit for Tat strategy renders individuals

vulnerable to being suckered on the first exchange (causing them to lose to selfish strategies one-on-on), Tit for Tat induces players to cut their losses quickly against those who exploit them, while reaping the benefits of cooperative exchanges with those who treat them right. Tit for Tat can be considered a somewhat moral strategy because it induces individuals to behave altruistically on the first exchange and fairly (quid pro quo) on subsequent exchanges. However, Tit for Tat would not be considered a highly moral strategy by people who believe it is right to forgive those who trespass against us, because Tit for Tat induces people to get even with those who treat them selfishly. In one of the first computer-based simulations of the evolution of social strategies, run by Robert Axelrod and his colleagues at the Institute of Public Policy Studies, University of Michigan in collaboration with the late William Hamilton, from Oxford University, Tit for Tat came out on top, defeating all other strategies against which it competed, including a purely selfish strategy called "All D" (always defect). This outcome excited many evolutionary theorists because it demonstrated that at least one cooperative strategy could defeat purely selfish strategies and evolve, which implied that we could inherit a capacity to behave in moral ways (Axelrod and Hamilton 1981).

At first, it seemed that Tit for Tat would reign supreme over all other strategies. However, follow-up studies revealed that this was not the case. As you might expect, Tit for Tat did well in populations replete with cooperative strategies; however, it did poorly in populations replete with selfish strategies. In addition to being vulnerable to exploitation on the first exchange, the Tit for Tat strategy was susceptible to getting locked into mutually defeating selfish-selfish exchanges. Let's say, for example, that two Tit for Tat strategists are happily enjoying the benefits of cooperative exchanges, but one of them makes a mistake or behaves selfishly in a moment of weakness. Obeying the Tit for Tat decision-rule, the other player returns selfishness with selfishness, which induces the first player to behave selfishly, giving rise to an endless string of self-defeating exchanges – a blood feud. Game theorists came to discover that conditionally cooperative strategies that enable players to break out of such self-defeating exchanges without opening themselves up to ongoing exploitation – strategies such as "Generous Tit for Tat" and "Tit for two Tats" – could defeat Tit for Tat. A hopeful aspect of these findings is that all of the strategies that defeated Tit for Tat were "nicer" than Tit for Tat – that is to say, they were more forgiving and induced players to make more altruistic and cooperative choices (Ridley 1996).

3.3.5 Paths to the Evolution of Cooperation

The key to explaining the evolution of goodness lies in explaining how cooperative social strategies that benefit others at immediate costs to actors can pay off genetically in the end, or how such strategies could have evolved even though they were maladaptive. Evolutionary theorists have identified five ways in which conditionally cooperative strategies could evolve. In all five cases, the question to keep in mind is

how these strategies help individuals propagate the genes that dispose them to behave in cooperative ways. To begin with, cooperative strategies such as Tit for Tat can evolve when they produce return benefits by inducing recipients to pay donors back (*direct reciprocity*). Second, cooperative strategies can evolve when they produce return benefits indirectly, by inducing third parties who are willing to cooperate to select cooperative individuals as exchange partners (*indirect reciprocity*). Third, cooperative strategies can evolve when individuals suffer the survival costs of helping others in order to increase their chances of propagating their genes by mating (*sexual selection*). Fourth, cooperative strategies can evolve when they increase the biological success of kin who share copies of cooperators' genes (*kin selection*). And finally, cooperative strategies can evolve when they increase the biological welfare of groups that contain members who share copies of cooperators' genes (*group selection*). In all five cases, the immediate costs that individuals suffer from helping others pay off genetically in the end. Inasmuch as conditionally cooperative strategies equate to conditionally moral strategies, they render those who invoke them conditionally good by nature. (Cooperative strategies also can evolve through cultural selection, but I will not discuss this process here.)

3.4 Psychological Sources of Goodness

Biologists focus on the evolution of behavioral strategies; evolutionary psychologists focus on the ways in which the evolved mental mechanisms that give rise to these strategies are designed. Lay people base their attributions of goodness primarily on the psychological sources of behaviors. When we observe others doing good things, we ask ourselves what kinds of motives are driving their behaviors – what they are trying to achieve. To deem a behavior moral, it must be viewed as stemming from a moral motive and being aimed at doing good. For example, if we think that a charitable act stemmed from sympathy for the disadvantaged and was aimed at improving their lot, we would be inclined to consider it moral, but if we thought that it was aimed at establishing dominance, impressing others, or currying favor we would be inclined to consider it selfish.

Mental mechanisms that give rise to love, guilt, gratitude, forgiveness, empathy and other "moral emotions" have evolved in our species. Other animals, especially other primates, display precursors of these emotions. The mental mechanisms that regulate these emotions reside in the older parts of our brains. As our brains evolved, we acquired higher-order mental mechanisms that enabled us to engage in uniquely-human forms of cognition such as perspective-taking and moral reasoning. These mechanisms, which reside in the outer layers of our cerebral cortex, normally become increasingly sophisticated as our brains develop throughout the lifespan. Early-evolved and later-evolved mechanisms participate in the activation and regulation of moral motives.

3.5 Moral Emotions

The function of moral emotions is to motivate individuals to behave in ways that enable them to reap the biological and genetic benefits of cooperation. Early humans who inherited psychological mechanisms that disposed them to experience moral emotions fared better than those who did not inherit these mechanisms. For example, early humans who inherited psychological mechanisms that induced them to love their offspring, siblings, and mates were more likely to propagate their genes through kin selection and sexual selection than those who did not. Groups that contained members who inherited mechanisms that disposed them to experience feelings of solidarity and loyalty fared better than more selfish groups in inter-group conflicts. Moral emotions such as gratitude, guilt, and righteous indignation paid off biologically and genetically by motivating our early ancestors to behave in ways that upheld biologically beneficial systems of social exchange. Feelings of forgiveness motivated early humans to repair broken social relations. A sense of justice evolved because it motivated our ancestors to uphold the cooperative social orders of their groups. Moral emotions are a significant psychological source of goodness.

3.5.1 Empathy

Among the suite of emotions that have been classified as moral, empathy and its cousin, sympathy, have received the most attention from social scientists. Neuroscientists have traced empathic reactions to "mirror neurons." When people observe others performing acts, these mirror neurons fire in exactly the same way as when they perform the acts themselves (Decety 2005). Theory and research on empathy illustrates the ways in which early-evolved and later-evolved mental mechanisms interact to motivate people to do good. As expressed by De Waal (2006), "empathy covers all the ways in which one individual's emotional state affects another's, with simple mechanisms at its core and more complex mechanisms and perspective-taking abilities at its outer layers. Because of the layered nature of the capacities involved, we speak of the Russian doll model, in which higher cognitive levels of empathy build upon a firm, hard-wired basis" (p. 11). Each "layer" gives rise to a different kind of empathic experience, or a different form of empathy. The more recently-evolved and advanced the level, the more altruistic the motives it engenders.

The Evolution of Empathy It seems plausible that the psychological mechanisms that endowed early humans with a capacity for empathy originated though kin selection. Early humans who inherited mechanisms that induced them to share the feelings of their offspring and other blood relatives propagated the genes that guided the creation of these mechanisms by helping those with whom they shared genes. As these mental mechanisms evolved, they came to be activated by individuals other than kin. As expressed by De Waal (2006), "the empathic response is amplified by

similarity, familiarity, social closeness, and positive experience with the other…. In human studies subjects empathize with a confederate's pleasure or distress if they perceive the relationship as cooperative." De Waal cites research showing that "seeing the pain of a cooperative confederate activates pain-related brain areas, but seeing the pain of an unfair confederate activates reward-related areas, at least in men." De Waal concludes that, "the empathy mechanism is biased the way evolutionary theory would predict. Empathy is (a) activated in relation to those with whom one has a close or positive relationship, and (b) suppressed, or even turned into Schadenfreude, in relation to strangers and defectors" (p. 16).

It is tempting to view our tendency to empathize with and assist people who share characteristics of our kin as a misfiring of mental mechanisms that evolved through kin selection, and this may, in part, be the case. However, in addition, such "overgeneralizations" probably benefited our early human ancestors by producing several return benefits. Early humans who helped non-kin with whom they empathized may well have been rewarded biologically and genetically by enhancing their reputations, by increasing their chances of mating, by upholding the groups on which they were dependent, by inducing recipients to reciprocate, and so on.

De Waal (2006) asserts that the first type of empathy to evolve, which constitutes the primitive core in humans and some other animals, induces us to experience feelings of *personal distress* when we are exposed to the distress of others – a type of emotional contagion. When we are in this state, we do not differentiate ourselves from those with whom we are empathizing. We are motivated to help others in order to relieve our own feelings of distress, which renders the motivation selfish in nature.

At "the next evolutionary step," which De Waal (2006) labels *cognitive empathy*, "emotional contagion is combined with appraisal of the other's situation and attempts to understand the cause of the other's emotions" (p. 9). Put another way, at this level, individuals attempt to take the perspective of those who are experiencing distress and understand it from their point of view. Cognitive empathy motivates individuals to engage in sympathetic behaviors such as consoling those who have been harmed. De Waal cites evidence demonstrating that consolation is common in humans and apes (and interestingly, in some large-brained birds), but virtually nonexistent in monkeys.

De Waal goes on to suggest that with expanded brain evolution, humans and a few other species acquired increasingly sophisticated perspective-taking abilities, which endowed them with the capacity to experience increasingly other-oriented forms of empathy. According to De Waal (2006), the essential cognitive ability that endows us with a capacity for the highest level of empathy – *empathic perspective taking* – is the ability to distinguish ourselves from others and to understand how others are feeling on their own terms: "For an individual to move beyond being sensitive to others toward an explicit other-orientation requires a shift in perspective. The emotional state induced in oneself by the other now needs to be attributed to the other instead of the self. A heightened self-identity allows a subject to relate to the object's emotional state without losing sight of the actual source of this state" (p. 9). De Waal (2006) cites evidence that apes, humans, elephants, and dolphins are

able to recognize themselves in mirrors, that this ability is correlated with perspective-taking abilities in humans, and that animals that possess self-recognition abilities engage in "targeted helping," defined as "help that is fine-tuned to another's specific situation and goals" (p. 9).

Neuroscience research on empathy has supported De Waal's assertion that empathic reactions in humans are produced by an interaction between primitive types of emotional contagion and more advanced forms of cognition. Based on a review of research in the area, the neuroscientist Decety (2005) concluded that "cognitive processes that exert a top-down control on [primitive empathic emotional reactions] are mediated by specific subregions of the prefrontal cortex" that help people distinguish actions that they produce from actions produced by others (p. 153).

The Development of Empathy De Waal accounts for the evolution of empathy in our species in terms of the increasing sophistication of cognitive abilities mediated by the evolution of the brain. The human brain also expands as children develop. Psychologists have advanced models of the development of empathy that are complementary to De Waal's evolutionary model. Consider, for example, the model advanced by Martin Hoffman (2000).

Hoffman (2000) acknowledges that empathy stems from evolved dispositions, then goes on to describe four phases in its growth in children that are defined in large part by the expansion and refinement of perspective-taking abilities. In the first phase, infants experience empathic reactions as "global distress," triggered, for example, by the cries of other infants. In the second phase, infants display egocentric empathic reactions that motivate them to respond to others' distress by engaging in behaviors that make them (but not those with whom they are empathizing) feel better. For example, they might hug their teddy bear. In the third phase, children's empathic reactions are evoked by interpretations of the situations that others experience, and children make more finely tuned emotional attributions, realizing that others' thoughts and feelings may differ from their own. In the final, most advanced, phase (not discussed by De Waal), those who develop sophisticated perspective-taking abilities acquire the ability to understand that others' reactions are affected by life experiences that go beyond the immediate situations at hand. Such people are able to experience empathy for disadvantaged groups or classes of people that they have never observed directly.

3.5.2 Empathy and Altruism

Empathizing with others in distress motivates us to help them – there is no question about that. However, there is some question about the extent to which such helping behaviors are aimed at relieving the suffering of victims, as opposed to improving the welfare of recipients, and therefore how altruistic they are. Daniel Batson, from the University of Kansas, has investigated this issue most extensively. Batson

(1991, 1998) launched his research program on what he called "the empathy-altruism hypothesis" by demonstrating that we may react in two quite different ways when we are exposed to the suffering of others. First, in concert with De Waal and Hoffman, Batson suggested that we may experience *personal distress* that we can allay in a variety of ways, such as looking away, leaving the scene, or helping the suffering person. Batson conceded that when we help others in order to relieve feelings of personal distress, our behavior is selfish in nature. However, argued Batson (2000), observing others who are feeling bad also may evoke "an other-oriented emotional response…[such as] empathy, sympathy, compassion, etc." (which he labeled *empathic concern*) that engenders "a motivational state with the … goal of increasing another's welfare" (pp. 207–8). Batson argued that empathic concern motivates us to help others as an end in itself, as opposed to helping them instrumentally in order to relieve our own vicariously experienced personal distress or to achieve other selfish goals.

Batson reasoned that if, as cynics claim, people who empathize with victims help them in order to reduce their personal distress (that is to say, if the goal they are attempting to achieve is to make themselves feel better), then they should not help victims when they are able to reduce their personal distress in less costly ways, such as leaving the scene. Batson and his colleagues designed experiments to test this hypothesis and found that participants who empathized with victims (but not participants who did not empathize with them) chose to help them even when given opportunities to reduce their distress in other ways. Batson concluded that empathy (but not personal distress) engenders altruistic motives.

However, Batson's findings did not satisfy some of his more cynical colleagues, who argued that the participants in Batson's experiments who seemed to be behaving altruistically could have been attempting to achieve selfish goals other than reducing their personal distress. For example, they could have been trying to avoid feeling guilty; they could have been trying to improve their mood; they could have been trying to make a good impression on the experimenter, and so on. Batson and his colleagues conducted some 30 experiments that evaluated the possibility that participants who experienced empathy with victims helped them in order to advance their own interests. Consider three examples. To evaluate the cynical claim that people help those with whom they empathize in order to gain social approval or to avoid disapproval, Batson and his colleagues determined whether people who empathized with victims were more likely to help in public than they were to help in private. To evaluate the claim that those who empathize with victims help them in order to avoid self-censure and guilt, Batson and his colleagues determined whether participants they had induced to empathize helped victims when they were offered personally and socially acceptable reasons and justifications for not helping. To evaluate the claim that those in empathic states help others in order to feel good about themselves, Batson and his colleagues assessed the mood of empathizing participants after they learned that third parties helped victims.

Batson and his colleagues were remarkably successful at disconfirming competing hypotheses and demonstrating that people choose to help those with whom they empathize even when they are prevented from using their helping behaviors to

achieve selfish goals, and even when they are offered more direct and less costly ways of achieving selfish goals than helping others. These investigators also found that when people who empathized with others in need were prevented from helping them, they felt bad. Citing a review of the literature on prosocial behavior by the psychologists Piliavin and Charng (1990), Batson (1998) concluded that a "paradigm shift" is occurring in psychology "away from the earlier position that behavior that appears to be altruistic must, under closer scrutiny, be revealed as reflecting egoistic motives. Rather, theory and data now being advanced are more compatible with the view that true altruism – acting with the goal of benefiting another – does exist and is a part of human nature" (p. 27). Natural goodness.

It is important to note that Batson and his colleagues did not claim that people do not ever help others for selfish reasons. Even in Batson's research, participants who were not induced to empathize with victims helped others in order to pursue selfish goals such as reducing their personal distress, gaining social approval, and improving their mood. Batson claimed only that a particular set of mental mechanisms – those that mediate advanced levels of empathy – engender altruistic motives. Note also that it is not possible to prove conclusively that participants in Batson's studies who empathized with victims were not driven to help them by some hidden selfish motive activated by Batson's methods of evoking empathy. And finally, note that Batson acknowledged that empathy does not always engender moral motives; it can also motivate people to behave in immoral (e.g., nepotistic and unjust) ways.

3.6 The Evolution and Development of Moral Reasoning

The Russian Doll model that De Waal invoked to explain the evolution of empathy can also be invoked to account for the evolution of moral reasoning. Moral reasoning probably originated as a tool used by early humans to advance their interests in strategic interactions with members of their groups. "You should be good to me because it will pay off for you in the end …." As humans evolved larger brains, their capacity to invoke increasingly sophisticated forms of moral reasoning would have increased in much the same manner it does in children as they develop. According to cognitive-developmental theorists such as Kohlberg (1984), moral reasoning is the central source of goodness in people. The better we become at moral reasoning, the greater our capacity to make fair and impartial moral decisions, and the more highly developed our sense of justice. However, viewing moral reasoning as a tool that evolved to help early humans advance their biological and genetic interests raises questions about its status as a source of goodness.

Do people use moral reasoning to reach impartial moral decisions, or do they use it to advance their own selfish interests? On the one hand, there is a great deal of evidence that most people are inclined to process information in ways that favor them, their relatives, and members of their in-groups (Pyszczynski and Greenberg 1987). For example, people tend to hold others to higher moral standards than they apply to themselves, and people are more likely to argue that others should sacrifice

their interests for their sake than that they should sacrifice their interests for others' sake. People tend to overestimate how much they deserve (their rights) and underestimate how much they owe (their duties), while underestimating how much others deserve and overestimating how much others owe (Greenberg and Cohen 1982). People tend to overvalue the contributions that they make to others, while undervaluing the contributions that others make to them. And people often use moral judgments to justify their immoral acts and to deflect responsibility onto others (Haidt 2001), diminishing their transgressions and excusing their misdeeds, while exaggerating the blameworthiness of their adversaries' immoral behavior (Krebs and Laird 1998).

Fortunately, however, there are several antidotes to self-serving biases in moral judgment. To begin with, people may constrain their self-serving biases in order to achieve the long-term benefits of moral strategies. In *The descent of man*, Charles Darwin (1874) suggested that early humans' capacity to reason enabled them to understand that it is in their long-term interest to forgo immediate gratification in order to uphold the interests of their communities and to figure out which customs are best equipped to help them improve their welfare: "As the reasoning powers and foresight of the members [of early human groups] became improved, each man would soon learn from experience that if he aided his fellow-men, he would commonly receive aid in return" (p. 127). On this line of thought, the capacity to reason facilitated the evolution of morality by enabling early humans to figure out that investing in their groups and behaving in moral ways paid off better than more selfish alternatives in the end – a form of enlightened self-interest.

In a similar vein, people may constrain their self-serving biases during the process of moral negotiations with others in order to avoid stalemates and reach win-win deals, and in addition, people may persuade themselves to abide by moral rules and principles in the process of attempting to persuade others. As expressed by Singer (1981), "ethical reasoning, once begun, pushes against our initially limited ethical horizons, leading us always toward a more universal point of view." "If I claim that what I do is right, while what you do is wrong, I must give some reason other than the fact that my action benefits me (or my kin, or my village) while your action benefits you (or your kin or your village)" (pp. 118–9). In a similar vein, Bloom (2004) asserted "once a creature is smart enough, impartiality – and an appreciation of moral codes such as the Golden Rule – will emerge as a consequence of this smartness" (p. 139). In a book entitled *The Better Angels of our Nature*, Pinker (2011) adduced a great deal of evidence in support of the conclusion that advances in our ability to reason accompanied by cultural developments such as mass media have mediated significant increases in the goodness of the human race over the past several thousand years.

3.7 Summary

Evolutionary theorists who have concluded that, on the laws of evolution, we (and all other animals) must be selfish by nature are not referring to the kind of selfishness that we consider immoral. The question that arises from an evolutionary analysis of morality is how behavioral strategies that we consider moral fared against behavioral strategies that we consider immoral. Did early human ancestors who treated their fellows right fare better than those who wronged them?

To understand the evolution of goodness, we must understand the adaptive functions that moral traits served in early human environments. Along with many other theorists, I argue that the central function of morality was (and still is) to uphold systems of cooperation that advanced the adaptive interests of members of groups. In large part, moral behaviors equate to cooperative behaviors. Evolutionary theorists have explained how a variety of conditionally cooperative strategies could evolve. Such strategies render us conditionally good (and conditionally bad) by nature.

A suite of psychological mechanisms has evolved to motivate us to behave in cooperative and moral ways. It is helpful to view these mechanisms in a Russian doll manner, with those that give rise to primitive emotions such as love and empathy at the core and those that give rise to higher-order cognitive processes such as perspective-taking and moral reasoning in the outer layers. Although perspective-taking and moral reasoning abilities probably originally evolved to help early humans prevail over members of their groups in strategic social interactions, there are several ways in which they can produce impartial moral judgments. Perspective-taking and moral reasoning abilities normally increase in sophistication as we develop, and they affect the ways in which we experience moral emotions, which increases our capacity for goodness. Cultural changes that render us increasingly civilized contribute to this process. Although we have not yet reached the summit of the mountain of goodness – indeed, it could be argued that we have a long way to go – we have made a great deal of progress.

References

Alexander, R.D. 1985. A Biological Interpretation of Moral Systems. *Zygon* 20: 3–20.

Axelrod, R., and W.D. Hamilton. 1981. The Evolution of Cooperation. *Science* 211: 1390–1396.

Batson, C.D. 1991. *The Altruism Question: Toward a Social-psychological Answer*. Hillsdale: Erlbaum.

———. [4]1998. Altruism and Prosocial Behavior. In *The Handbook of Social Psychology*, ed. D.T. Gilbert, S.T. Fiske, and G. Lindzey, 282–315. Boston: McGraw-Hill.

———. 2000. Unto Others: A Service… and a Disservice. *Journal of Consciousness Studies* 7: 207–210.

Bloom, P. 2004. *Descartes' Baby: How the Science of Child Development Explains What Makes us Human*. New York: Basic Books.

Brown, R. 1984. *Social Psychology*. 2nd ed. New York: The Free Press.

Campbell, D. 1978. On the Genetics of Altruism and the Counterhedonic Components of Human Nature. In *Altruism, Sympathy, and Helping*, ed. L. Wispe, 39–58. New York: Academic Press.
Curry, O.S. 2016. Morality as Cooperation: A Problem-centered Approach. In *The evolution of Morality*, ed. T.K. Shackelford and R.D. Hansen, 27–52. Cham: Springer.
Darwin, C. 1874. *The Descent of Man and Selection in Relation to Sex*. New York: Rand, McNally & Co.
Dawkins, R. 1989. *The Selfish Gene*. Oxford: Oxford University Press.
de Waal, F.B.M. 2006. *Primates and Philosophers: How Morality Evolved*. Princeton: Princeton University Press.
———. 2008. Putting the Altruism Back in Altruism. *Annual Review of Psychology* 59: 279–300.
Decety, J. 2005. Perspective Taking as the Royal Avenue to Empathy. In *Other Minds: How Humans Bridge the Divide between Self and Other*, ed. B.F. Malle and S.D. Hodges, 143–157. New York: Guilford Press.
Ellis, B.J. 1998. The Partner-specific Investment Inventory: An Evolutionary Approach to Individual Differences in Investment. *Journal of Personality* 66: 383–442.
Flack, J.C., and F.B.M. de Waal. 2000. 'Any Animal Whatever': Darwinian Building Blocks of Morality in Monkeys and Apes. *Journal of Consciousness Studies* 7: 1–29.
Greenberg, J., and R.L. Cohen. 1982. *Equity and Justice in Social Behavior*. New York: Academic Press.
Haidt, J. 2001. The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review* 108: 814–834.
Hoffman, M. 2000. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press.
Kohlberg, L. 1984. *Essays in Moral Development: The Psychology of Moral Development*. Vol. Vol 2. New York: Harper & Row.
Krebs, D.L. 2011. *The Origins of Morality: An Evolutionary Account*. New York: Oxford University Press.
Krebs, D.L., and P. Laird. 1998. Judging Yourself as You Judge Others: Perspective-taking, Moral Development, and Exculpation. *Journal of Adult Development* 5: 1–12.
Piliavin, J.A., and H.-W. Charng. 1990. Altruism: A Review of Recent Theory and Research. *Annual Review of Sociology* 16: 27–65.
Pinker, S. 2011. *The Better Angels of our Nature*. New York: Viking.
Pyszczynski, T., and J. Greenberg. 1987. Toward an Integration of Cognitive and Motivational Perspectives on Social Inference: A Biased Hypothesis-testing Model. *Advances in Experimental Social Psychology* 20: 297–340.
Rawls, J. [2]1999. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Ridley, M. 1996. *The Origins or Virtue: Human Instincts and the Evolution of Cooperation*. New York: Viking.
Singer, P. 1981. *The Expanding Circle: Ethics and Sociobiology*. New York: Farrar, Straus & Giroux.
Sober, E., and D.S. Wilson. 2000. Summary of Unto Others: The Evolution and Psychology of Unselfish Behavior. In *Evolutionary Origins of Morality: Cross-disciplinary Perspectives*, ed. L.D. Katz, 185–206. Thorverton: Imprint Academic.
Williams, G.C. 1989. A Sociobiological Expansion of "Evolution and Ethics.". In *Evolution and Ethics*, ed. J.G. Paradis and G.C. Williams, 179–214. Princeton: Princeton University Press.